I.A.3.9 SUMEX Staff Publications

The following are publications for the SUMEX staff and include papers describing the SUMEX-AIM resource and on-going research as well as documentation of system and program developments. Many of the publications documenting SUMEX-AIM community research are from the individual collaborating projects and are detailed in their respective reports (see Section II on page 81). Publications for the AGE and AI Handbook core research projects are given there.

[1] Carhart, R.E., Johnson, S.M., Smith, D.H., Buchanan, B.G., Dromey, R.G., and Lederberg, J, Networking and a Collaborative Research Community: A Case Study Using the DENDRAL Programs, ACS Symposium Series, Number 19, Computer Networking and Chemistry, Peter Lykos (Editor), 1975.

[2] Levinthal, E.C., Carhart, R.E., Johnson, S.M., and Lederberg, J., When Computers Talk to Computers, Industrial Research, November 1975

[3] Wilcox, C. R., MAINSAIL - A Machine-Independent Programming System, Proceedings of the DEC Users Society, Vol. 2, No. 4, Spring 1976.

[4] Wilcox, Clark R., The MAINSAIL Project: Developing Tools for Software Portability, Proceedings, Computer Application in Medical Care, October, 1977, pp. 76-83.

[5] Lederberg, J. L., Digital Communications and the Conduct of Science: The New Literacy, Proc. IEEE, Vol. 66, No. 11, Nov 1978.

[6] Wilcox, C. R., Jirak, G. A., and Dageforde, M. L., MAINSAIL - Language Manual, Stanford University Computer Science Report STAN-CS-80-791 (1980).

[7] Wilcox, C. R., Jirak, G. A., and Dageforde, M. L., MAINSAIL - Implementation Overview, Stanford University Computer Science Report STAN-CS-80-792 (1980).

Mr. Clark Wilcox also chaired the session on "Languages for Portability" at the DECUS DECsystem10 Spring '76 Symposium.

In addition, a substantial continuing effort has gone into developing, upgrading, and extending documentation about the SUMEX-AIM resource, the SUMEX-TENEX system, and the many subsystems available to users. These efforts include a number of major documents (such as SOS, PUB, TENEX-SAIL, and MAINSAIL manuals) as well as a much larger number of document upgrades, user information and introductory notes, an ARPANET Resource Handbook entry, and policy guidelines.

I.A.3.10 Future Plans

Our plans for the next grant year are based on the council-approved plans for our five-year renewal that began in August 1980. In addition to the specific plans for next grant year (discussed in some earlier sections too), we include a summary of the overall objectives for this five-year period to serve as a background. Near and long term objectives and plans for individual collaborating projects are discussed in Section II beginning on page 81.

Overall Goals

The goals of the SUMEX-AIM resource are long term in supporting basic research in artificial intelligence, applying these techniques to a broad range of biomedical problems, experimenting with communication technologies to promote scientific interchange, and developing better tools and facilities to carry on this research. Just as the tone of our renewal proposal derives from the continuing long-term research objectives of the SUMEX-AIM community, our approach derives from the methods and philosophy already established for the resource. We will continue to develop useful knowledge-based software tools for biomedical research based on innovative, yet accessible computing technologies.

For us it is important to make systems that work and are exportable. Hence, our approach is to integrate available state-of-the art hardware technology as a basis for the underlying software research and development necessary to support the AI work. SUMEX-AIM will retain its broad community orientation in choosing and implementing its resources. We will draw upon the expertise of on-going research efforts where possible and build on these where extensions or innovations are necessary. This orientation has proved to be an effective way to build the current facility and community.

We have built ties to a broad computer science community; have brought the results of their work to the AIM users; and have exported results of our own work. This broader community is particularly active in developing technological tools in the form of new machine architectures, language support, and interactive modalities.

Toward a More Distributed Resource

The initial model for SUMEX as a centralized resource was based on the high cost of powerful computing facilities and not being able to duplicate them readily. This role is evolving, though, with the introduction of more compact and inexpensive computing technology. Our future goals are guided by community needs for more computing capacity and improved tools to build more effective expert systems and to test operational versions of AI programs in real-world settings. In order to meet these needs, we must take advantage of a range of newly developing machine architectures and systems. As a result, SUMEX-AIM will become more a distributed community resource with heterogeneous computing facilities tethered to each other through communications media. Many of these machines will be located physically near the projects or biomedical scientists using them.

We have actively supported proposals from the more mature AIM projects for additional computing facilities tailored to their particular needs and designed to free the main SUMEX resource for new, developing applications projects. To date, the Rutgers resource has acquired a DEC 2060 facility, part of which is allocated for AIM usage; the "Simulation of Cognitive Processes" project has acquired a VAX which supports their needs; and the CADUCEUS project has acquired a VAX to support experimental clinical testing of their program. Our future plans anticipate an even broader diversification of computing resources to meet the need of the AIM community.

The Continuing Role of SUMEX-Central

Even with more distributed computing resources, the central resource will continue to play an important role as a communication crossroads, as a research group devoted to integrating the new software and hardware technologies to meet the needs of medical AI applications, as a spawning ground for new application projects, and as a base for local AI projects. A key challenge will be to maintain the scientific community ties that grew naturally out of the previous colocation within a central facility.

Summary of Five-year Objectives

The long term objectives of the SUMEX-AIM resource nucleus during the follow-on five year period are summarized below. These are broken into three categories; resource operations, training and education, and core research.

Resource Operations

1) Maintain the vitality of the AIM community. We will continue to encourage and explore new applications of AI to biomedical research and improve mechanisms for inter- and intra-group collaborations and communications. While AI is our defining theme, we may entertain exceptional applications justified by some other unique feature of SUMEX-AIM essential for important biomedical research. To minimize administrative barriers to the community-oriented goals of SUMEX-AIM and to direct our resources toward purely scientific goals, we plan to retain the current user funding arrangements for projects working on SUMEX facilities. User projects will fund their own manpower and local needs; will actively contribute their special expertise to the SUMEX-AIM community; and will receive an allocation of computing resources under the control of the AIM management committees. There will be no "fee for service" charges for community members. We will also continue to exploit community expertise and sharing in software development; and to facilitate more effective information sharing among projects.

2) Provide effective computational support for AIM community goals. We will continue to expand support for artificial intelligence research and new applications work; to develop new computational tools to support more mature projects; and to facilitate testing and research dissemination of nearly operational programs. We will continue to operate and develop the existing central facility as the nucleus of the resource. We will acquire additional equipment to meet developing community needs for more capacity, larger program address spaces, and improved interactive facilities. New computing hardware technologies becoming available now and in the next few years will play a key role in these developments and we expect to take the lead in this community for adapting these new tools to biomedical AI needs.

3) Provide effective and geographically accessible communication facilities to the SUMEX-AIM community for effective remote collaborations, communications among distributed computing nodes, and experimental testing of AI programs. We will retain the current ARPANET and TYMNET connections for at least the near term and will actively explore other advantageous connections to new communications networks and to dedicated links.

Training and Education

1) Assist new and established projects in the effective use of the SUMEX-AIM resource. Collaborating projects continue to be responsible for the development and dissemination of their own AI programs but the resource staff will provide general support and will work to make resource goals and AI systems known and available to appropriate biomedical scientists. We will continue to exploit particular areas of expertise within the community for developing pilot efforts in new application areas.

2) Continue to allocate "collaborative linkage" funds to qualifying new and pilot projects to provide for communications and terminal support pending formal approval and funding of their projects. These funds are allocated in cooperation with the AIM Executive Committee reviews of prospective user projects.

3) Continue to support workshop activities including collaboration with the Rutgers Computers in Biomedicine resource on the AIM community workshop and with individual projects for more specialized workshops covering specific application areas or program dissemination.

Core Research

1) Continue to explore basic artificial intelligence research issues for knowledge acquisition, representation, and utilization; reasoning in the presence of uncertainty; strategy planning; and explanations of reasoning pathways with particular emphasis on biomedical applications. SUMEX core research funding is complementary to similar funding from other agencies and contributes to the long-standing interdisciplinary effort at Stanford in basic AI research and expert system design. We expect this work to provide the underpinnings for increasingly effective consultative programs in medicine and for more practical adaptations of this work within emerging microelectronic technologies.

2) Support community efforts to organize and generalize AI tools that have been developed in the context of individual application projects. This will include work to organize the present state-of-the-art in AI techniques through the AI Handbook effort and the development of practical software packages (e.g., AGE, EMYCIN, UNITS, and EXPERT) for the acquisition, representation, and utilization of knowledge in AI programs. The objective is to evolve a body of software tools that can be used to more easily build future knowledge-based systems and explore other biomedical AI applications.

Specific Plans for Year 10

Specific plans for the next grant year (10) are summarized in the paragraphs below. The directions and background for much of this work were given in earlier progress report sections and are not repeated in detail here.

Hardware Acquisition Plans

Our rationale for future hardware acquisitions is based on a balanced development of the SUMEX-AIM resource along the lines approved by council. As discussed in our progress report and supported by collaborating project reports, we have implemented an effective set of computing resources to support AI applications to biomedical research. Currently at the resource core are the dual KI-TENEX, 2020, shared VAX/UNIX, and Dolphin InterLisp systems, augmented by portions of the Rutgers and Stanford SCORE 2060 machines and the dedicated Caduceus VAX system at the University of Pittsburgh. These have provided an excellent set of tools for SUMEX-AIM development.

As the size of our community and the complexity of knowledge-based programs have increased, several issues have become crucial for the continued development and practical dissemination of AI programs. These have been thoroughly documented in our renewal proposal and previous reports and include the need for more computing capacity, the need for larger user virtual address spaces, and practical facilities to test and disseminate AI systems in the general biomedical community.

No single solution to these requirements for future development is available. We proposed and got peer approval to investigate a variety of machine architectures and approaches over the next grant period including new shared centralized systems, distributed single-user workstations, and improved communications tools to integrate these systems together effectively. Our approach is to integrate a heterogeneous set of network-connected hardware tools, some of which will be distributed through the user community. We will emphasize the development of system and application level software tools to allow effective use of these resources and continue to provide community leadership to encourage scientific communications.

Specific Hardware Plans for Year 10 -- Our equipment plan for year 10 is based on the council-approved grant award and previous experience gained with alternative systems.

Central Resource Upgrade: Having purchased the 5 InterLisp Dolphins last year, we are scheduled to upgrade the central resource this year. In our renewal application, this upgrade was tentatively planned to be a VAX system. However, based on our work during year 09 with using the shared VAX/UNIX system for AI Lisp development (see page 14), we feel the acquisition of additional VAX resources is not the most advantageous for the SUMEX-AIM community.

At the same time, the cost effectiveness of PDP-20 systems has been reemphasized. In addition to their demonstrated high performance and extensive software facilities, there is a growing suite of software that takes advantage of the recently available user version of extended addressing (including Lisp implementations). Thus, with the increasing isolation and expense of our local efforts to continue to develop and maintain the KI-10 system and the long-standing pressing need for more computing capacity in the SUMEX-AIM community, we are planning to upgrade the existing KI-10 system to a large DEC 2060 system. We will submit a plan for approval by the AIM Executive Committee, NIH, and council late this summer.

File Server Development: We began acquisition of initial components of a file server toward the end of year 09, including a VAX 11/750 processor and a starting configuration of non-DEC disks and tapes (see page 16). During year 10 we plan to add additional high capacity disk drives to this central server, accessible over the Ethernet, to help minimize the need for redundant large file systems on individual workstations and host machines.

Other Equipment: As part of the continued operation of the SUMEX-AIM resource, we also plan to buy additional required communications, Ethernet, terminal, interface, and test equipment to support community needs.

Continued Operation of Existing Hardware

The current SUMEX-AIM facilities represent a large existing investment. As indicated above, we plan to upgrade the KI-TENEX system to a 2060 TOPS-20 machine. We do not propose any substantial changes to the other hardware systems (2020, shared VAX, and Dolphins) and we expect them to continue to provide effective community support and serve as a communication nucleus for more distributed resources. The current file shortage will be remedied by access to the community file server, sharable and accessible via the Ethernet.

Communication Networks

Networks have been centrally important to the research goals of SUMEX-AIM and will become more so in the context of increasingly distributed computing. Communication will be crucial to maintain community scientific contacts, to facilitate shared system and software maintenance based on regional expertise, to allow necessary information flow and access at all levels, and to meet the technical requirements of shared equipment.

Long-Distance Connections -- We have had reasonable success at meeting the geographical needs of the community during the early phases of SUMEX-AIM through our ARPANET and TYMNET connections. These have allowed users from many locations within the United States and abroad to gain terminal access to the AIM resources and through ARPANET links to communicate much more voluminous file information. Since many of our users do not have ARPANET access privileges for technical or administrative reasons, a key problem impeding remote use has been the limited communications facilities (speed, file transfer, and terminal handling) offered currently by commercial networks. Commercial improvements are slow in coming but may be expected to solve the file transfer problem in the next few years. A number of vendors (AT&T, IBM, Xerox, etc.) have yet to announce commercially available facilities but TELENET is actively working in this direction. We plan to continue experimenting with improved facilities as offered by commercial or government sources in the next grant term. We have budgeted for continued TYMNET service and an additional amount annually for experimental network connections.

High-speed interactive terminal support will continue to be a problem since one cannot expect to serve 1200-9600 baud terminals effectively over shared long-distance trunk lines with gross capacities of only 9600-19200 baud. We feel this is a problem that is best solved by distributed machines able to effectively support terminal interactions locally and coupled to other AIM machines and facilities through network or telephonic links. As new machine resources are introduced into the community, we will allocate budgeted funds with Executive Committee advice to assure effective communication links.

Local Intermachine Connections -- A key feature of our plans for future computing facilities is the support of a heterogeneous processing environment that takes advantage of newly available technology and shared equipment resources. The "glue" that links these systems together is a high speed local network. We have chosen Ethernet and the Xerox PUP [9, 12] protocols for these interconnections. This choice was based on the early availability of that technology and the economics of using already developed TENEX and other server software. We expect the Ethernet system to continue to meet our technical needs for the coming grant term and we plan to continue to use it. We are working closely with other groups here at Stanford and elsewhere to share hardware interface and software designs wherever possible.

Our goals are to continue development of the SUMEX-AIM network system including completion of the TIP server, development of a shared printer server for various existing serial printers, implementation of the necessary interface for the new 2060, improvement of the file service facilities between hosts and workstations, and beginning development of an Ethernet TYMNET gateway.

Resource Software

We will continue to maintain the existing system, language, and utility support software on our systems at the most current release levels, including up-to-date documentation. We will also be extending the facilities available to users where appropriate, drawing upon other community developments where possible. We rely heavily on the needs of the user community to direct system software development efforts.

Our plans for upgrading the SUMEX-AIM KI-TENEX system will entail substantial effort to assure smooth user transition to the new system. We will also have to reorient users as different versions of common utility programs run on the TOPS-20 system than under TENEX. Our goals will be to continue to derive as much software as possible from the TOPS-20, UNIX, and workstation user communities but we expect to have to do considerable work to adapt them to our biomedical AI needs.

Within the AIM community we expect to serve as a center for software sharing between various distributed computing nodes. This will include contributing locally developed programs, distributing those derived from elsewhere in the community, maintaining up-to-date information on subsystems available, and assisting in software maintenance.

Community Management

We plan to retain the current management structure that has worked so well, including the newly created GENET Executive Committee which oversees applications in molecular biology (see page 73). We will continue to work closely with the management committees to recruit the additional high quality projects which can be accommodated and to evolve resource allocation policies which appropriately reflect assigned priorities and project needs. We expect the Executive and Advisory Committees to play an increasingly important role in advising on priorities for facility evolution and on-going community development planning in addition to their recruitment efforts. The composition of the Executive committee will grow as needed to assure representation of major user groups and medical and computer science applications areas. The Advisory Group membership rotates regularly and spans both medical and computer science research expertise. We expect to maintain this policy.

We will continue to make information available about the various projects both inside and outside of the community and thereby promote the kinds of exchanges exemplified earlier and made possible by network facilities.

The AIM workshops under the Rutgers resource have served a valuable function in bringing community members and prospective users together. We will continue to support this effort. This summer the AIM workshop will be held at the University of Pittsburgh in conjunction with the American Association of Artificial Intelligence conference. We will continue to assist community participation and provide a computing base for workshop demonstrations and communications. We will also assist individual projects in organizing more specialized workshops as we have done for the DENDRAL and AGE projects.

We plan to continue indefinitely our present policy of non-monetary allocation control. We recognize, of course, that this accentuates our responsibility for the careful selection of projects with high scientific and community merit.

Training and Education Plans

We have an on-going commitment, within the constraints of our staff size, to provide effective user assistance, to maintain high quality documentation of the evolving software support on the SUMEX-AIM system, and to provide software help facilities such as the HELP and Bulletin Board systems. These latter aids are an effective way to assist resource users in staying informed about system and community developments and solving access problems. We plan to take an active role in encouraging the development and dissemination of community knowledge resources such as the AI Handbook, up-to-date bibliographic sources, and developing knowledge bases. Since much of our community is geographically remote from our machine, these on-line aids are indispensable for self help. We will continue to provide on-line personal assistance to users within the capacity of available staff through the SNDMSG and LINK facilities.

We budget funds to continue the "collaborative linkage" support initiated during the first term of the SUMEX-AIM grant. These funds are allocated under Executive Committee authorization for terminal and communications support to help get new users and pilot projects started.

Core Research Plans

SUMEX core research includes both basic AI research and development of community tools useful for building expert systems. Expert systems are symbolic problem solving programs capable of expert-level performance, in which domain-specific knowledge is represented and used in an understandable line of reasoning. The programs can be used as problem solving assistants or tutors, but also serve as excellent vehicles for research on representation and control of diverse forms of knowledge. MYCIN is one of the best examples.

Because the main issues of building expert systems are coincident with general issues in AI, we appreciate the difficulty of proposing to "solve" basic problems. However, we do propose to build working programs that demonstrate the feasibility of our ideas within well defined limits. By investigating the nature of expert reasoning within computer programs, the process is "demystified". Ultimately, the construction of such programs becomes itself a well-understood technical craft.

The foundation of all of our core research work is expert knowledge: its acquisition from practitioners, its accommodation into the existing knowledge bases, its explanation, and its use to solve problems. Continued work on these topics provides new techniques and mechanisms for the design and construction of knowledge-based programs; experience gained from the actual construction of these systems then feeds back both (a) evaluative information on the ideas' utility and (b) reports of quite specific problems and the ways in which they have been overcome, which may suggest some more general method to be tried in other programs.

One of our long-range goals is to isolate AI techniques that are general, to determine the conditions for their use and to build up a knowledge base about AI techniques themselves. SUMEX resources are coordinated for this purpose with the multidisciplinary efforts of the Stanford Heuristic Programming Project (HPP). Under support from ARPA, NIH/NLM, ONR, NSF, and private funding, the HPP conducts research on five key scientific problem areas, as well as a host of subsidiary issues [1]:

1) Knowledge Representation - How shall the knowledge necessary for expert-level performance be represented for computer use? How can one achieve flexibility in adding and changing knowledge in the continuous development of a knowledge base? Are there uniform representations for the diverse kinds of specialized knowledge needed in all domains?

2) Knowledge Utilization - What designs are available for the inference procedure to be used by an expert system? How can the control structure be simple enough to be understandable and yet sophisticated enough for high performance? How can strategy knowledge be used effectively?

3) Knowledge Acquisition - How can the model of expertise in a field of work be systematically acquired for computer use? If it is true that the power of an expert system is primarily a function of the quality and completeness of the knowledge base, then this is the critical "bottleneck" problem of expert systems research.

4) Explanation - How can the knowledge base and the line of reasoning used in solving a particular problem be explained to users? What constitutes an acceptable explanation for each class of users?

5) Tool Construction - What kinds of software packages can be constructed that will facilitate the implementation of expert systems, not only by the research community but also by various user communities?

Artificial Intelligence is largely an empirical science. We explore questions such as these by designing and building programs that incorporate plausible answers. Then we try to determine the strengths and weaknesses of the answers by experimenting with perturbations of the systems and extrapolations of them into new problem areas. The test of success in this endeavor is whether the next generation of system builders finds the questions relevant and the answers applicable to reduce the effort of building complex reasoning programs.

I.B Highlights

I.B.1 DENDRAL - Computers in Chemistry

The DENDRAL project at Stanford University is one of the founding projects in the SUMEX-AIM community. The roots of its pioneering research can be traced even farther back to the mid-1960's and the interdisciplinary collaboration of principal investigators Profs. Joshua Lederberg (Genetics), Edward Feigenbaum (Computer Science), and Carl Djerassi (Chemistry). DENDRAL is a resource-related research project concerned with symbolic computational techniques to assist with determining topological and stereochemical aspects of organic molecular structures based on information obtained from physical and chemical methods. The project has also worked energetically to disseminate these techniques to a wide community of biochemists.

DENDRAL has gained fame as one of the "classical" research efforts in artificial intelligence and it is worth highlighting its seminal contributions to computer science, chemistry, and the SUMEX-AIM community. DENDRAL was among the very earliest AI projects to point out the now well-established tenet that "knowledge is power". It's effectiveness as a consultant in structural chemistry problems does not derive from a powerful, universal, problem-solving engine. Rather it comes from the large amount of domain specific knowledge it has about the manipulation of molecular structure graphs, mass spectrometry, NMR spectrometry, and other areas of biochemical expertise. Most current AI projects now focus heavily on the issues DENDRAL raised about knowledge acquisition, representation, and utilization. The paradigm of knowledge-based expert systems promoted by DENDRAL research has received wide recognition in the AI literature.

From the viewpoint of the SUMEX-AIM and biochemistry communities, DENDRAL has been a path-finding project as well. It has traversed the full cycle of development from concept to "breadboard" to prototype to operational program, all the while pursuing innovative research on new application areas. The staff of the project has been dedicated to making the DENDRAL programs useful not only in their own laboratory but also to a broader community of biochemists. The significant impact of these programs on the biochemical community can be measured by the large number of productive collaborations that have developed between the DENDRAL staff and other research groups and the more than 20 sites running DENDRAL software.

DENDRAL has worked closely with the SUMEX staff in devising new and effective means for disseminating complex software systems through the various phases of project development. A wide range of approaches has been explored including network access to programs on SUMEX; graduate courses and hands-on workshops at Stanford; reprogramming parts of the system for export to other machine environments and network-accessible systems; and creation of an industrial affiliates group and licensing the programs through Stanford for commercial development, marketing, and maintenance. Similar approaches are now being used effectively by other projects in the community.

I.B.2 CADUCEUS - Diagnostic Assistance in Internal Medicine

For the past 10 years, Dr. Jack Myers has been collaborating with computer scientist Prof. Harry Pople at the University of Pittsburgh on the development of a computer program to assist with complex diagnoses in the broad field of internal medicine. Major achievements of this research using SUMEX-AIM have been the development and experimentation with an early diagnostic program called INTERNIST-I, the on-going design and implementation of an improved program called CADUCEUS, and the development of an extensive medical knowledge base covering many aspects of internal medicine.

INTERNIST-I has been used to analyze many hundreds of difficult diagnostic problems. In most cases, the program has performed at the level of a skilled internist and is one of the most spectacular examples of artificial intelligence applications in biomedicine. It is capable of handling multiple diagnoses of related or independent diseases in a given patient and has been exercised on complex cases published in medical journals (particularly Case Records of the Massachusetts General Hospital, in the New England Journal of Medicine), CPC's, and unusual problems of patients in the University of Pittsburgh Medical Center. Its failures have highlighted areas needing continued research and improvement leading to the refinements in the CADUCEUS program.

The medical knowledge base has continued to grow both in the incorporation of new diseases and the modification of diseases already profiled so as to include recent advances in medical knowledge. The knowledge base now includes 536 individual disease profiles, 3,679 manifestations of disease, and about 1,500 "links" or interrelationships among diseases as well as a myriad of miscellaneous pieces of information which are essential for the correct operation of the system. Twenty-two new diseases have been profiled during the past year and the pediatrics and neurology portions of the knowledge base have continued to grow as well. Approximately 200 additional diseases remain to be programmed and renewed effort in this direction is now being mounted

Future development plans include continued refinement and testing of the extensive medical knowledge base; completing implementation of the improved program, CADUCEUS; initial clinical testing of CADUCEUS within the University of Pittsburgh and at other university health centers; and adaptation of the program and knowledge base for medical education.

I.B.3 GENET - Computing for Molecular Biology

Since early 1980, the MOLGEN project investigators at Stanford have made a new set of computing tools available to a national community of molecular biologists through a guest facility called GENET on the SUMEX-AIM resource. This experimental sub-community was started by Drs. Kedes, Brutlag, and Friedland in order to broaden MOLGEN's base of scientist collaborators at institutions other than Stanford and to explore the use of a SUMEX-like resource to disseminate sophisticated software tools to a generally computer-naive community. The enthusiastic response to our very limited announcement of this facility has continued to grow so that we have been obliged to place severe restrictions on the scope of services provided.

Since late 1980, we have limited the number of simultaneous GENET users to 2 and have established only a common shared directory among all the users with a very limited file space allocation. Even with such a relatively inhospitable working environment, in 1980-81 the GENET community consumed CPU resources at roughly half the rate of the largest AI research project on SUMEX. This year GENET used almost 50% more resources than the largest AI project! Currently GENET comprises several hundred users from over 60 research institutions.

We have offered three main programs to assist molecular genetics users: SEQ, a DNA-RNA sequence analysis program; MAP, a program that assists in the construction of restriction maps from restriction enzyme digest data; and MAPPER (written and maintained by William Pearson from Johns Hopkins University), a simplified version of the MOLGEN MAP program that is somewhat more efficient than the MOLGEN version. Some of the other more sophisticated programs being developed by MOLGEN research efforts have not been offered because they are not ready for novice users. In addition, the GENET users have had access to the SUMEX-AIM programs for electronic messaging, text editing, file searching, etc.

We have succeeded in the goals set out for GENET to the point that the response has been frankly overwhelming. The SUMEX resource is unable to provide the operational support needed for GENET computing without seriously diminishing the system capacity needed for other AI research. Efforts to expand our own capacity to provide for both communities have not been successful. Over the past year, however, commercial alternatives to the SUMEX GENET service have come into existence. We are therefore in the process of restructuring the GENET activity on SUMEX to reemphasize the computer science research aspect of these applications in molecular biology. A separate GENET Executive Committee has been established to help administer this part of the resource. Strictly production use of the current GENET programs on SUMEX will be discontinued shortly and such users will be assisted in finding alternatives.

I.B.4 Handbook of Artificial Intelligence

The AI Handbook is a compendium of knowledge about the field of Artificial Intelligence. It has been edited by Avron Barr, Paul Cohen, and Edward Feigenbaum at Stanford University, with textual contributions from students and investigators at several other research institutions across the nation. The scope of the work is broad: Hundreds of articles cover most of the important ideas, techniques, and systems developed during 26 years of research in AI. Each short article is a description written for non-AI specialists and students of AI. Additional articles serve as overviews which discuss the various approaches within a subfield, the issues, and the problems.

There is no comparable resource for AI researchers and other scientists and technologists who need access to descriptions of AI techniques and concepts. The research literature in AI is not very accessible. And the elementary textbooks are not nearly broad enough in scope to be useful to a scientist working primarily in another discipline who wants to do something requiring knowledge of AI. Furthermore, we feel that some of the Overview articles are the best critical discussions of activity in the field available anywhere.

The major work of this project is now finished. The Handbook material was completed in April 1982, and has been published in three volumes -- over 1500 pages. Both the first and second volumes of the Handbook have been selected by the Library of Science Book Club as main selections. The chapters also are appearing as Stanford Computer Science Department Technical Reports available through the National Technical Information Service. Work continues on developing a convenient mechanism for on-line access to the Handbook material. When that software is completed, the Handbook text will be available for browsing by the SUMEX-AIM community.

The AI Handbook Project was undertaken as a core activity by SUMEX in the spirit of community building that is the fundamental concern of the resource. We feel that the organization and propagation of this kind of information to the AIM community, as well as to other fields where AI is being applied, is a valuable service that we are uniquely qualified to support. The Handbook articles are a good example of community collaboration using the SUMEX-AIM communication facilities to prepare, review, and disseminate this reference work on AI techniques. All of our authors and reviewers have had access to these files via the network facilities and use the document-editing and formatting programs available at SUMEX. This relatively small investment of resources has resulted in what we feel is a seminal publication in the field of AI -- of particular value to researchers, like those in the AIM community, who want quick access to AI ideas and techniques for application in other areas.

The three volume Handbook was published by Wm. Kaufmann, Inc., Los Altos, California. Volume I appeared in May 1981, and Volumes II and III appeared in June 1982.

I.B.5 Professional Workstations - A First Step

One of the most exciting prospects of the coming decade in computing is the development of professional workstations. These may have a profound impact on biomedicine by serving as the vehicle for the practical export of expert artificial intelligence systems into the hands of the physician, chemist, biologist, engineer, or other users. The essential concept of these machines is simple -- a large-address-space, high-performance, microcoded processor tuned to the Lisp programming language; a high-speed, high-resolution graphics display station for convenient human interactions; a local disk; a local area network interface; and well-human-engineered software all compactly packaged and priced for the professional office environment. These systems differ from the popularly available microprocessor systems in terms of their specialization and performance to support large-address Lisp programs that form the core of AI research and in their use of large-format, bit-mapped graphics displays.

Several research Lisp machines have become available during this past year. SUMEX has acquired and integrated 5 Xerox Dolphin InterLisp workstations into the computing environment at Stanford so research projects could begin work on the important software issues facing us in order to exploit personal computing environments for AI systems. Much systems work on Ethernet services has been done to make the Dolphins work smoothly and effectively in our existing facility. This has been shared widely with other research groups outside of Stanford in order to broaden and assist the growing user community. Research has also begun on moving the ONCOCIN, MOLGEN, AGE, and GUIDON systems to run on the Dolphins with promising early results. Shortly, one of the SUMEX Dolphins will be installed at Rutgers to enable researchers there to begin exploring its potential for biomedical AI systems.

Still newer research Lisp machines that extend current capabilities in performance, Lisp dialects and software systems supported, and cost-effectiveness will become available in the next years. By the mid-1980's, we expect a much broader availability of such systems for use by biomedical professionals in general.

I.C Administrative Changes

There have been no significant administrative changes in the SUMEX-AIM resource this past year.

I.D Resource Management and Allocation

The mission of SUMEX-AIM, locally and nationally, entails both the recruitment of appropriate research projects interested in medical AI applications and the catalysis of interactions among these groups and the broader medical community. User projects are separately funded and autonomous in their management. They are selected for access to SUMEX on the basis of their computer and biomedical scientific merits as well as their commitment to the community goals of SUMEX. Currently active projects span a broad range of application areas such as clinical diagnostic consultation, molecular biochemistry, molecular genetics, belief systems modeling, mental function modeling, and instrument data interpretation (descriptions of the individual collaborative projects are in Section II beginning on page 81).

I.D.1 Management Committees

Since the SUMEX-AIM project is a multilateral undertaking by its very nature, we have created several management committees to assist in administering the various portions of the SUMEX resource. As defined in the SUMEX-AIM management plan adopted at the time the initial resource grant was awarded, the available facility capacity is allocated 40% to Stanford Medical School projects, 40% to national projects, and 20% to common system development and related functions. Within the Stanford aliquot, Prof. Feigenbaum and BRP have established an advisory committee to assist in selecting and allocating resources among projects appropriate to the SUMEX mission. The current membership of this committee is listed in Appendix D.

For the national community, two committees serve complementary functions. An Executive Committee oversees the operations of the resource as related to national users and makes the final decisions on authorizing admission for new projects and revalidating continued access for existing projects. It also establishes policies for resource allocation and approves plans for resource development and augmentation within the national portion of SUMEX (e.g., hardware upgrades, significant new development projects, etc.). The Executive Committee oversees the planning and implementation of the AIM Workshop series currently implemented under Prof. S. Amarel of Rutgers University and assures coordination with other AIM activities as well. The committee will play a key role in assessing the possible need for additional future AIM community computing resources and in deciding the optimal placement and management of such facilities. The current membership of the Executive committee is listed in Appendix D.

Reporting to the Executive Committee, an Advisory Group represents the interests of medical and computer science research relevant to AIM goals. The Advisory Group serves several functions in advising the Executive Committee; 1) recruiting appropriate medical/computer science projects, 2) reviewing and recommending priorities for allocation of resource capacity to specific projects based on scientific quality and medical relevance, and 3) recommending policies and development goals for the resource. The current Advisory Group membership is given in Appendix D.

This past spring, the AIM Executive Committee authorized the establishment of a GENET Executive Committee to oversee the management of the now significant portion of the national portion of SUMEX allocated to molecular biology. The committee was established by NIH/BRP for this purpose and to advise BRP more generally in the management and long term planning of molecular biology support. The membership of the GENET Executive Committee is given in Appendix D. Its first task is to redirect the GENET resources from production use of the DNA sequence mapping programs that have grown so popular toward new computer science research for molecular biology is now underway.

These committees have actively functioned in support of the resource. Except for the meetings held during the AIM workshops, the committees have "met" by messages, net-mail, and telephone conference owing to the size of the groups and to save the time and expense of personal travel to meet face to face. The telephone meetings, in conjunction with terminal access to related text materials, have served quite well in accomplishing the agenda business and facilitate greatly the arrangement of meetings. Other solicitations of advice requiring review of sizable written proposals are done by mail.

We will continue to work with the management committees to recruit the additional high quality projects which can be accommodated and to evolve resource allocation policies which appropriately reflect assigned priorities and project needs. We will continue to make information available about the various projects both inside and outside of the community and thereby promote the kinds of exchanges exemplified earlier and made possible by network facilities.

I.D.2 New Project Recruiting

The SUMEX-AIM resource has been announced through a variety of media as well as by correspondence, contacts of NIH-BRP with a variety of prospective grantees who use computers, and contacts by our own staff and committee members. The number of formal projects that have been admitted to SUMEX has more than trebled since the start of the project to a current total of 7 national AIM projects and 8 Stanford projects. Others are working tentatively as pilot projects or are under review.

We have prepared a variety of materials for the new user ranging from general information such as is contained in a SUMEX-AIM overview brochure to more detailed information and guidelines for determining whether a user project is appropriate for the SUMEX-AIM resource. A questionnaire is available to assist users seriously considering applying for access to

SUMEX-AIM. Pilot project categories have been established both within the Stanford and national aliquots of the facility capacity to assist and encourage new projects in formulating possible AIM proposals and pending their application for funding support. Pilot projects are approved for access for limited periods of time after preliminary review by the Stanford or AIM Advisory Group as appropriate to the origin of the project. More recently we have established a category of "associates" representing both US and international groups working on their own machines on problems relevant to the SUMEX-AIM community and seeking more direct communication with SUMEX-AIM investigators. Associate community members are approved by the AIM Executive Community.

These contacts have sometimes done much more than provide support for already formulated programs. For example, Prof. Feigenbaum's group at Stanford previously initiated a major collaborative effort with Dr. Osborn's group at the Institutes of Medical Sciences in San Francisco. This project in "Pulmonary Function Monitoring and Ventilator Management - PUFF/VM" (see page 194) originated as a pilot request to use MLAB in a small way for modeling. Subsequently the AI potentialities of this domain were recognized by Feigenbaum, Nii, and Osborn and a joint proposal was submitted to and funded by NIH. Similarly, Prof. Feigenbaum and Ms. Nii have worked directly with Profs. Kintsch and Polson at the University of Colorado, introducing them to the newly developed AGE package for use in formulating their program on modeling aspects of human cognition.

A list of the fully authorized projects currently comprising the SUMEX-AIM community can be found with brief abstracts in Appendix C on page 277. More detailed descriptions of collaborative project activities can be found in Section II.

As an additional aid to new projects or collaborators with existing projects, we provide a limited amount of funds for use to support terminals and communications needs of users without access to such equipment. We are currently providing support for 6 terminals and 4 modems for users as well as a leased line between Stanford and the University of California at Santa Cruz for the Chemical Synthesis project.

I.D.3 Stanford Community Building

The Stanford community has undertaken several internal efforts to encourage interactions and sharing between the projects centered here. Professor Feigenbaum organized a project with the goal of assembling a handbook of AI concepts, techniques, and current state-of-the-art. This project has had enthusiastic support from the students and recently completed publication of volumes II and III of the 3 volume handbook set (see page 92 for more details).

Weekly informal lunch meetings (SIGLUNCH) are also held between community members to discuss general AI topics, concerns and progress of individual projects, or system problems as appropriate. In addition, presentations from a substantial number of outside speakers are invited.

I.D.4 Existing Project Reviews

We have conducted a continuing careful review of on-going SUMEX-AIM projects to maintain a high scientific quality and relevance to our medical AI goals and to maximize the resources available for newly developing applications projects. At meetings of the AIM Advisory Group and Executive Committee this past year, all the national AIM projects were reviewed. These groups recommended continued access for most formal projects on the system. However, they recommended that the GENET effort be guided toward developing new computer science applications in molecular biology rather than continuing production use of the current DNA sequence analysis programs.

I.D.5 Resource Allocation Policies

As the SUMEX facility has become increasingly loaded, a number of diverse and conflicting demands have arisen which require controlled allocation of critical facility resources (file space and central processor time). Our policy for file space management starts with an allocation of file storage, defined for each authorized project in conjunction with the management committees. This allocation is divided among project members in any way desired by the individual principal investigators. System enforcement of project allocations is carried out each week. As the weekly file dump is done, if the aggregate space in use by a project is over its allocation, files are archived from user directories over allocation until the project is within its limits.

We have also implemented effective system scheduling controls to attempt to maintain the 40:40:20 balance in terms of CPU utilization and to avoid system and user inefficiencies during overload conditions. The initial complement of user projects justifying the SUMEX resource was centered to a large extent at Stanford. Since then, a substantial relative growth in the number of national projects was realized and in practice the 40:40 split between Stanford and non-Stanford projects is fairly well realized (see Figure 8 on page 37 and the tables of recent project usage on page 40).

Our job scheduling controls bias the allocation of CPU time based on percent time consumed relative to the time allocated with the 40:40:20 community split. The controls are "soft" however in that they do not waste computer cycles if users below their allocated percentages are not on the system to consume the cycles. In early years, the operating disparity in CPU use reflected a substantial difference in demand between the Stanford community and the developing national projects, rather than inequity of access. For example, the Stanford utilization is spread over a large part of the 24-hour cycle, while national-AIM users tended to be more sensitive to local prime-time constraints. (The 3-hour time zone phase shift across the continent is of substantial help in load balancing.) During peak times under the overload control system reported previously, the Stanford community experienced mutual contentions and delays while the AIM group has relatively open access to the system.

This disparity in usage has disappeared in recent years though with the growth of the national user community and we enabled overload controls for the national community as well. For the present, we propose to continue our policy of "soft" allocation enforcement for the fair split of resource capacity.

Our system also categorizes users in terms of access privileges. These comprise fully authorized users, pilot projects, associates, guests, and network visitors in descending order of system capabilities. We want to encourage bona fide medical and health research people to experiment with the various programs available with a minimum of red tape while not allowing unauthenticated users to bypass the advisory group screening procedures by coming on as guests. So far we have had relatively little abuse compared to what other network sites have experienced, perhaps on account of the personal attention that senior staff gives to the logon records, and to other security measures. However, the experience of most other computer managers behooves us to be cautious about being as wide open as might be preferred for informal service to pilot efforts and demonstrations. We will continue developing this mechanism in conjunction with management committee policy decisions.

We have also encouraged mature projects to apply for their own machine resources in order to preserve the SUMEX-AIM resource for research and development efforts and to support projects unable to justify their own machines. The Rutgers resource has their own machine, part of which is allocated for AIM use, and the CADUCEUS project has installed a VAX machine to support their planned development and program testing work. Also Profs. Lesgold and Greeno's "Simulation of Cognitive Processes" project has moved to their own local VAX.

I.E Dissemination Efforts

Throughout its existence, SUMEX-AIM has devoted substantial efforts toward disseminating information about its activities as a resource and about the work of individual collaborative projects. We continue to make many presentations at professional meetings, to provide services to demonstrate developed AI programs for interested groups and individuals, to welcome visitors, and to work in organizing workshops within the SUMEX-AIM community to introduce our work to collaborating professional communities. We have also spent substantial efforts in the past working with the Research Resources Information Center to produce the "Seeds of Artificial Intelligence" monograph and other publications and press articles to address a broader community of technical and lay people.

Of special note, in addition to these general dissemination activities, is the development of the GENET molecular biology community:

GENET Background

The MOLGEN project at Stanford (see page 128) has focused on applications of artificial intelligence and symbolic computation to the field of molecular biology. The research began in 1975 and is currently working under a three year grant renewal. In early 1980 it was realized that some of the systems developed by MOLGEN were of direct utility to many scientists in the domain. Accordingly, with the cooperation of the SUMEX-AIM staff and close coordination with the AIM Executive Committee, it was decided in February 1980 to provide a carefully limited guest service for the community use of such systems.

There were two major reasons for the establishment of this guest service, which took the form of the GENET account on SUMEX. The first was to broaden MOLGEN's base of scientist collaborators, to find molecular biologists at institutions other than Stanford who could contribute actively to our knowledge-based approach to problem solving. The second was to introduce a generally computer-naive community to the benefits of resource sharing provided by a system like SUMEX, with the hope of serving as a model for the dissemination of other AI software and possibly for an eventual resource for molecular biology.

We believe that we have succeeded in these two goals. Many of our GENET guests have become active collaborators in core MOLGEN research. These collaborators include Professor Allan Maxam at Harvard Medical School, Dr. Walter Goad at Los Alamos, Dr. Richard Roberts at Cold Spring Harbor, Dr. William Pearson at Johns Hopkins, Drs. Walter Bodmer, Julia Bodmer, and Robert Kamen at the Imperial Cancer Research Fund, Dr. Andrew Taylor at University of Oregon, and Dr. Dan Davison of SUNY-Stonybrook. We are also pleased by the numerous comments SUMEX has received from GENET users praising the user-sensitive nature of the resource, especially in comparison to typical university computer centers.

GENET has been important both for MOLGEN and for the national community of molecular biology. It has ensured a steady flow of ideas for the artificial intelligence research that is core to both the MOLGEN grant and the SUMEX-AIM mission. It has also provided a useful service to an international community that has not been readily available elsewhere.

GENET Community Management

Our decision to support the GENET guest experiment and our approach to doing so within the SUMEX-AIM resource has been reviewed and approved both by the AIM Executive Committee and by the Initial Review Group/National Advisory Research Resources Council in the course of the peer review of our latest SUMEX renewal application. We have tried to manage the GENET guest experiment in such a way that we maintain the "friendly" interface of the SUMEX-AIM resource for molecular biologists unfamiliar with computers while taking appropriate steps so that GENET usage does not detract from on-going AI research and so that we assure prudent administration SUMEX as an NIH-BRP resource. The key elements in our management approach include:

1) Controlled announcement of the GENET opportunity -- Beginning in February 1980, the availability of GENET services was announced, primarily by talks at professional conferences with accompanying program demonstrations. We decided against publishing "blanket" announcements in professional journals in order to maintain a very high standard of collaborator interest and scientific expertise within the limited group we could serve with available SUMEX resources.

2) Close coordination with the AIM Executive Committee -- We kept the AIM Executive Committee apprised of plans for the GENET experiment and of progress and growth of the community. At the August 1980 AIM Workshop meeting of the Executive Committee, Professor L. Kedes of the MOLGEN project made a presentation on the status of GENET. The Executive Committee approved continuation of the GENET service but because of the significant growth in the number of GENET users and their consumption of CPU resources, a limit of two simultaneous GENET jobs was placed on the community. The Executive Committee also approved the concept of a proposed Molecular Biology Computing Resource related to but separate from the existing SUMEX resource.

3) Careful control of GENET usage -- We have closely monitored the very rapid growth in GENET usage of SUMEX (see data below). With Executive Committee advice and in cooperation with the MOLGEN project personnel managing the GENET community, we have instituted several successively stringent controls on GENET users:

   a) All GENET users run out of the same directory so scheduler control limits are enforced to hold GENET usage as a whole down relative to that of AI research projects during heavy loads.

   b) The GENET directory has been intentionally limited in disk space allocation so that large numbers of files cannot be retained.

c) Starting in October 1980, a limit of two simultaneous logged-in GENET jobs was placed on the community.

d) Starting in December 1980, a policy statement was issued restricting GENET use to academic collaborators. MOLGEN project management informed industrial collaborators that they could no longer use the GENET facility and actively monitored adherence to this policy. Previously, valuable feedback had been obtained from a small group of industrial collaborators for MOLGEN AI program development. However, with the rapid growth of the highly competitive molecular genetics industry, there was no way we could adequately control industrial users consistent with SUMEX's status as a federally funded national resource. Thus, we decided to exclude them. In April 1981, we instituted a GENET user password checking system to further control community access, particularly in regard to industrial users.

4) Limited commitment of SUMEX staff resources -- The day to day management of the GENET community has been the responsibility of MOLGEN project personnel. SUMEX personnel have only contributed to developing system facilities to help manage GENET (guest and GENET password capabilities), assisted with technical communications problems, and advised in establishing GENET management policies consistent with AIM Executive Committee and SUMEX Principal Investigator resource policies. The total commitment of staff time has been on the order of 1-2 man-months.

5) Creation of the GENET Executive Committee -- With all of the previous limitations on the GENET community, its usage continued to grow to the point of dominating AI research use on the national side. Since previous efforts to establish a separate national resource for molecular biology were not funded and alternative services have started becoming available commercially, the AIM Executive Committee authorized establishing a GENET Executive Committee to manage the use of the GENET aliquot of SUMEX and to advise NIH/BRP more broadly on developing support for research in molecular biology. This committee has met and is now formulating a plan to remove the production use of SUMEX by GENET and to reallocate these resources to new computer science developments for the molecular biology community.

Scope of the GENET User Community

The GENET community consists of approximately 200-300 users from 63 research institutions. Of these users, approximately 35 are consistently active users. That is, they log in, run programs, and interact with the MOLGEN members on an almost daily basis. Many of these users have made valuable contributions to our work. About 100 others are frequent, but not regular users. They log in only when they have a major analysis task to perform, which seems to be on the order of once a month.